Mateo Alado

San Lorenzo, CA | (510) 543-1444 | aladomateo@gmail.com | linkedin.com/in/mateoalado | github.com/matesuu

Education

California State University, East Bay August 2024 – December 2026
Bachelor of Science in Computer Science *Hayward, CA*

- Distinctions: 2x AI Research Publications, HackHayward 2026 Winner, 3x Dean's List
- Coursework: Data Structures and Algorithms, Software Engineering, Operating Systems, AI, Statistics

Professional Experience

Software Engineer Intern (Agents) June 2026 – August 2026
Lawrence Berkeley National Laboratory *Berkeley, CA*

- Engineered a pipeline of 4 LangGraph agents for scientific literature ingestion and retrieval of 16,000+ publications
- Built a PostgreSQL-GraphQL API for data retrieval, cutting task completion time from 3+ minutes to <5 seconds
- Deployed agents on an HPC cluster using Slurm, scaling literature extraction and processing throughput by 70%
- Containerized deployment of agents and services with Docker, expanding access to 450+ ALS beamline scientists
- Resolved a request concurrency bottleneck by implementing a semaphore, increasing concurrent API calls by 25%
- Created 100+ end-to-end backend/agent tests with pytest, increasing total codebase test coverage 95%+

Software Research Engineer Intern January 2026 – May 2026
California State University, East Bay *Hayward, CA*

- Developed a FastAPI backend with context retrieval for data management, cutting request context size by 79.7%
- Built a script for automating data migration to a NoSQL database via cronjob, improving processing time by 60%
- Optimized a BFS adaptive graph traversal algorithm for retrieval, optimizing LLM response relevancy to 89%
- First-authored an IEEE paper on GenAI applications in academic research for the GAISS 2026 conference

Projects

Perplx - HackHayward 2026 Winner | *TypeScript, Bun, Cloudflare Workers, Cloudflare D1, MCP*

- Developed an AI coding agent with MCP tooling using the Perplexity API for context-aware software engineering
- Reduced cost per completed task to approximately $0.10 through auto compaction and prompt caching
- Built a session-sharing infrastructure using Cloudflare Workers and D1 enabling real-time collaborative sessions via shared URLs for 100+ concurrent users

Publication Management Platform | *Python, TypeScript, React, FastAPI, Neo4j, Supabase*

- Built and deployed a full-stack RAG research platform using React and FastAPI for 20+ users to query and manage 100+ research publications and datasets
- Implemented persistent session context retrieval and user data storage with Supabase and Neo4j, cutting end-to-end latency by 75%+
- Integrated LiteLLM to benchmark model accuracy and performance for 5 models across 3 different LLM providers

Visual Speech Classifier | *Python, PyTorch, NumPy, CNN, MediaPipe, OpenCV*

- Created a real-time lip-reading classifier and pipeline using MediaPipe Face Mesh and OpenCV webcam capture to classify speech patterns for hearing-impaired individuals
- Achieved a 94% validation accuracy across 24 speech classes through iterative CNN optimization and refinement

Skills

Languages: Python, TypeScript, JavaScript, Java, SQL, C, C++, Go
AI/ML: LLMs, AI Agents, LangGraph, LangChain, RAG, PyTorch
Backend: FastAPI, REST APIs, GraphQL, Node.js, Express
Data: PostgreSQL, MongoDB, NoSQL, Supabase, Neo4j
Frontend: React, React Native, Next.js, Tailwind
Cloud and Tools: Docker, GCP, AWS, Git, Linux, Slurm, Cursor, Claude Code
Concepts: Backend Development, Fullstack Development, Cloud Infrastructure, Artifical Intelligence, Machine Learning, CI/CD, DevOps, Automation, Agile